www.bsc.es

7th IICWG Data Assimilation and
PPP Ice Verification Workshop
ESA, Frascati, 5-7th April 2016

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Better observations, better forecasts?

F. Massonnet

O. Bellprat, V. Guemas, F. J. Doblas-Reyes

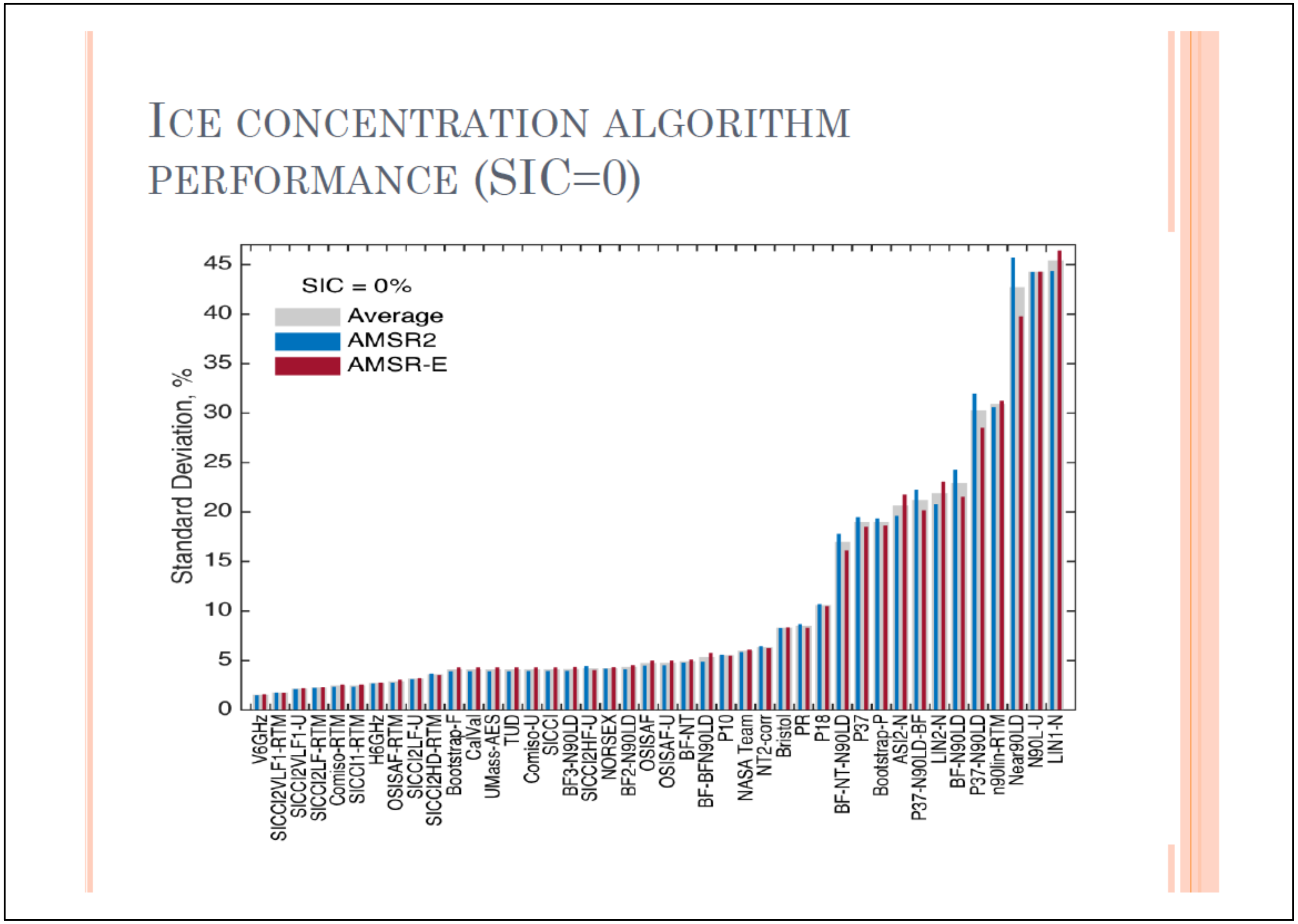
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

SPECS
Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

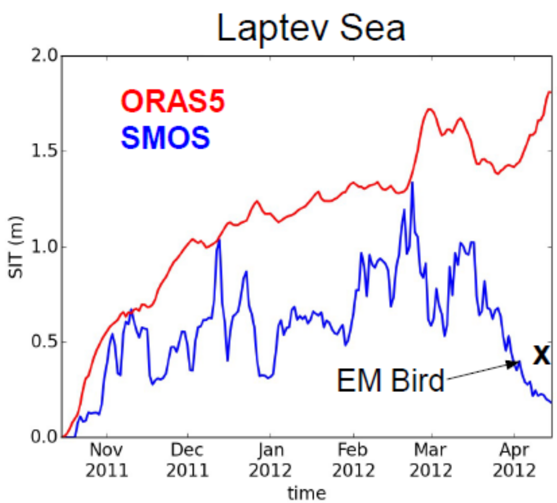
A variety of sea ice concentration products, a large variability in their quality

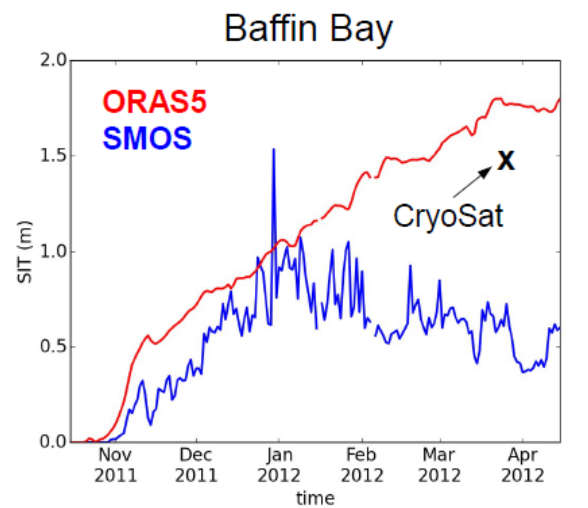


Courtesy L. T. Pedersen,
N. Ivanova and co-authors

Models and observations: which one is the reference? The frontier can be (very) thin

Winter 2011/2 in two selected locations





Independent data supports SMOS.

→ Model cannot simulate Polynya

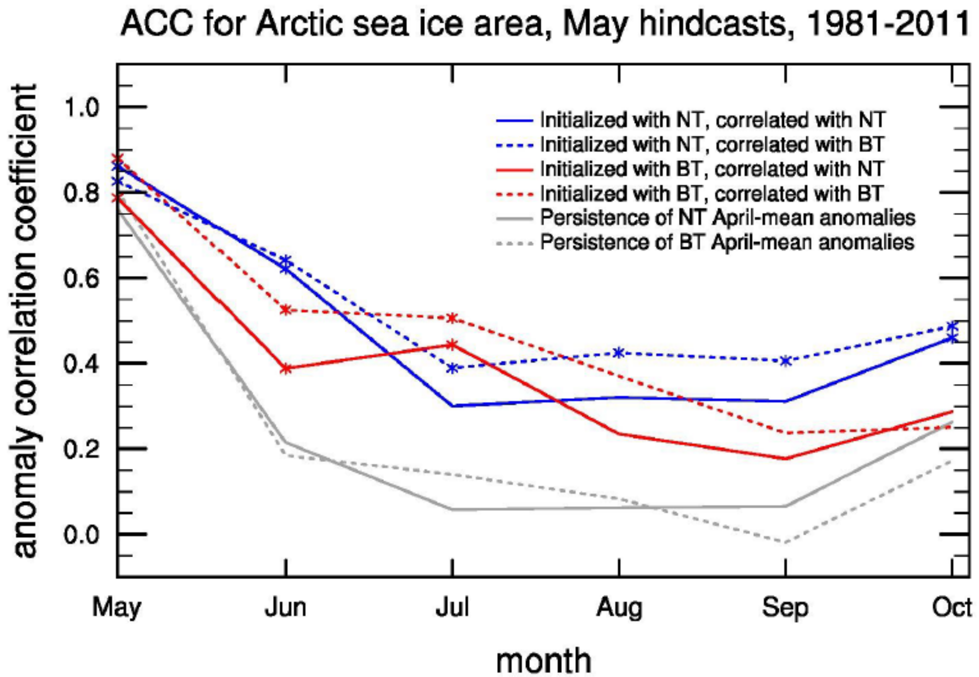
Independent data supports ORAS5.

→ SMOS assumptions on ice T, S, snow?
Both CryoSat + model wrong?

ECMWF

Courtesy S. Tietsche

ACC for Arctic sea-ice area

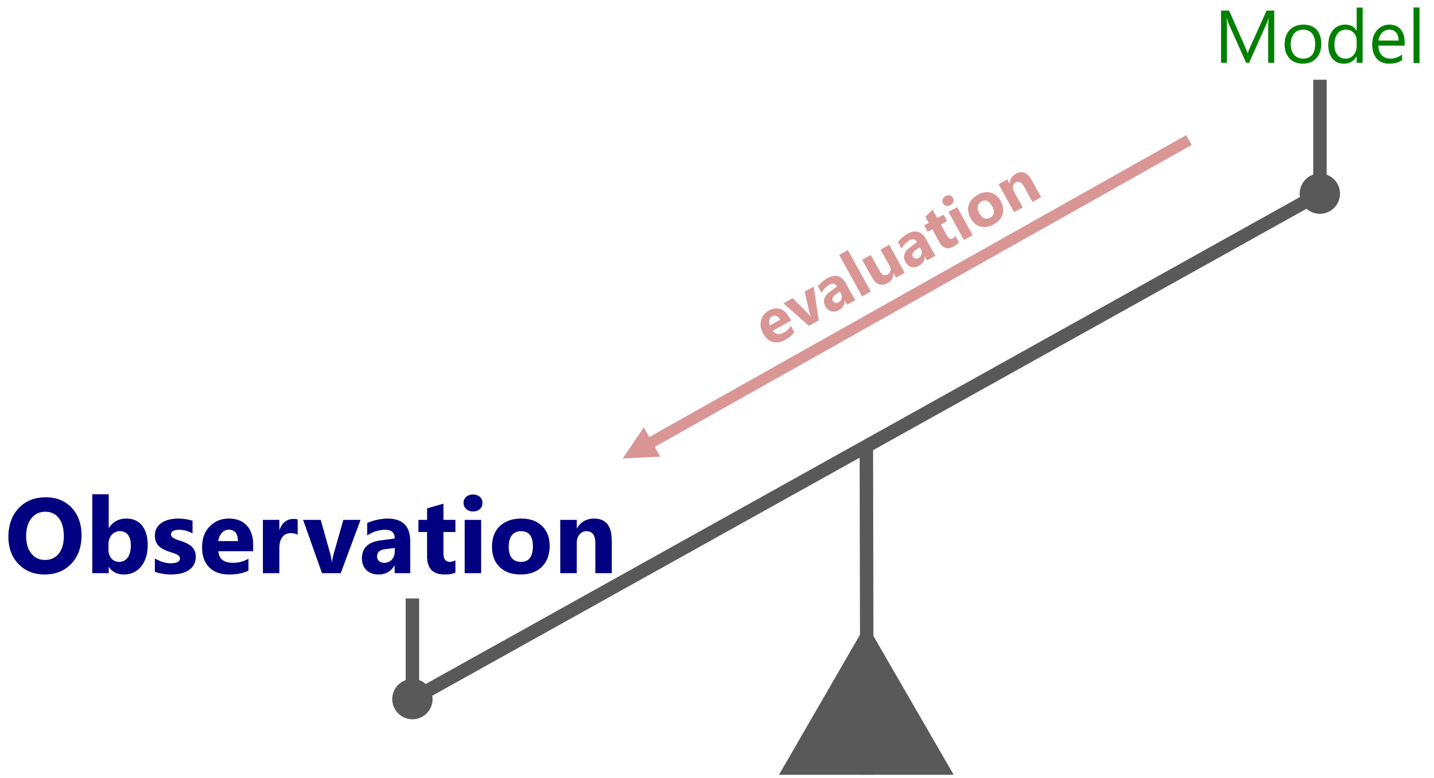


Initialization: Except for month 3 (July) NT initialization outperforms BT initialization

Reference: BT data produces larger ACC than NT data when used as reference

Courtesy F. Bunzel

My understanding
in 2010



There is a strong need for good sea ice observational products

Sea ice cover controls ocean-atmosphere heat fluxes
SIC and SIT products are used to drive atmosphere-only climate models

Tracing back sea ice model biases is important to improve models
Budget analyses e.g., Holland et al., Nature Geosci., 2012.

Observations are key in prediction/projection/attribution studies
e.g., Notz and Marotzke, GRL, 2012; Massonnet et al., The Cryosph. 2012

There is a strong need for good sea ice observational products

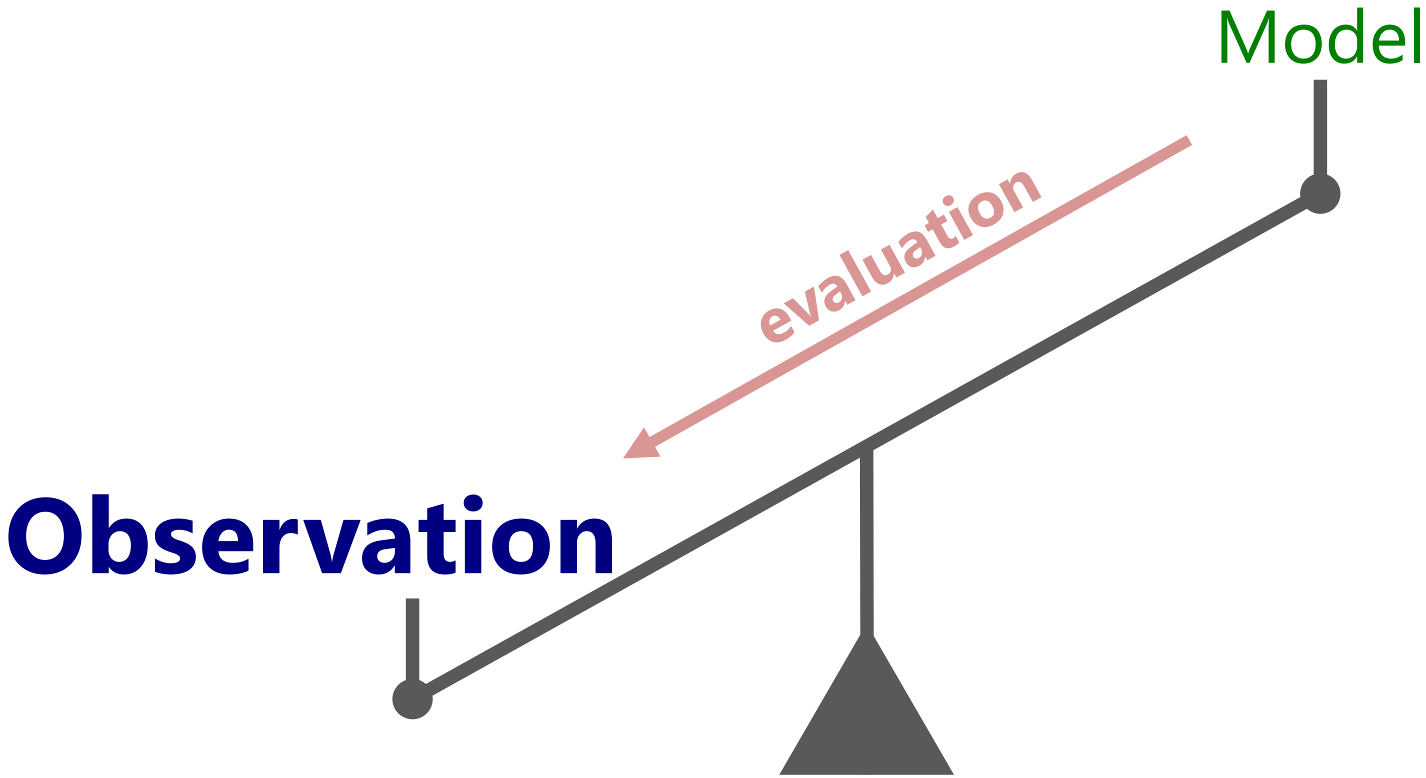
Sea ice cover controls ocean-atmosphere heat fluxes
SIC and SIT products are used to drive atmosphere-only climate models

Tracing back sea ice model biases is important to improve models
Budget analyses e.g., Holland et al., Nature Geosci., 2012

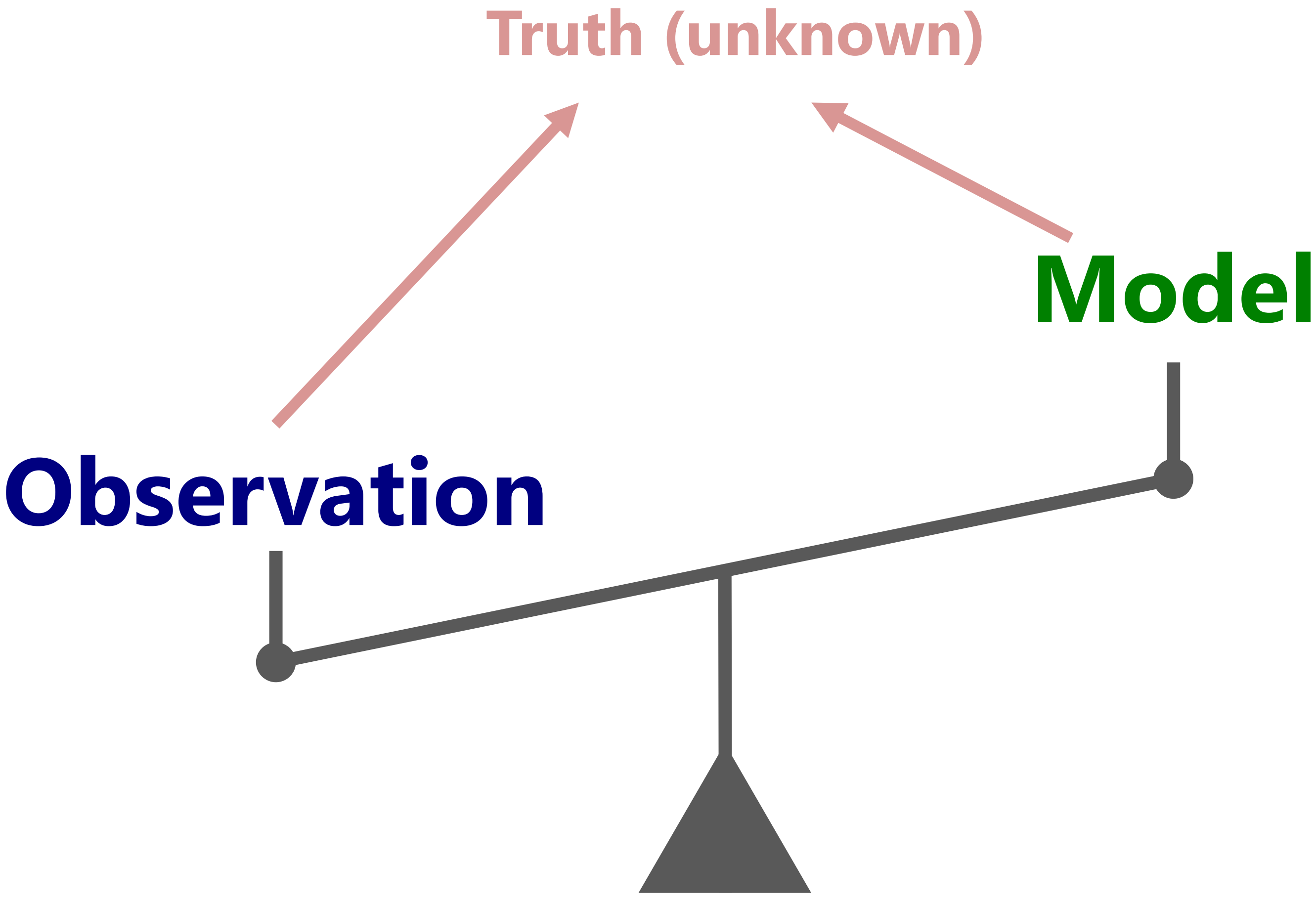
Observations are key in prediction/projection/attribution studies
e.g., Notz and Marotzke, GRL, 2012; Massonnet et al., The Cryosph. 2012

Sea ice observations are central for data assimilation
e.g., Massonnet et al., Oc. Modell., 2013

My understanding
in 2010



My understanding in 2013



There is a strong need for good sea ice observational products

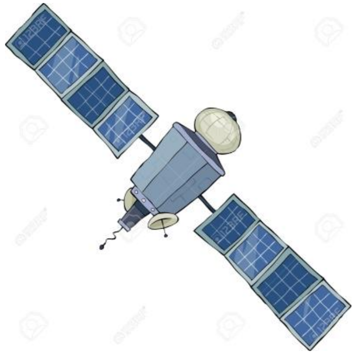
Sea ice cover controls ocean-atmosphere heat fluxes
SIC and SIT products are used to drive atmosphere-only climate models

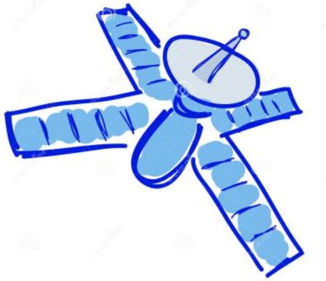
Tracing back sea ice model biases is important to improve models
Budget analyses e.g., Holland et al., Nature Geosci., 2012

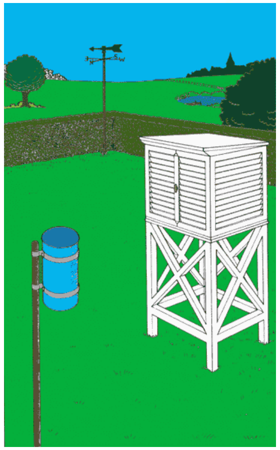
Observations are key in prediction/projection/attribution studies
e.g., Notz and Marotzke, GRL, 2012; Massonnet et al., The Cryosph. 2012

Sea ice observations are central for data assimilation
e.g., Massonnet et al., Oc. Modell., 2013

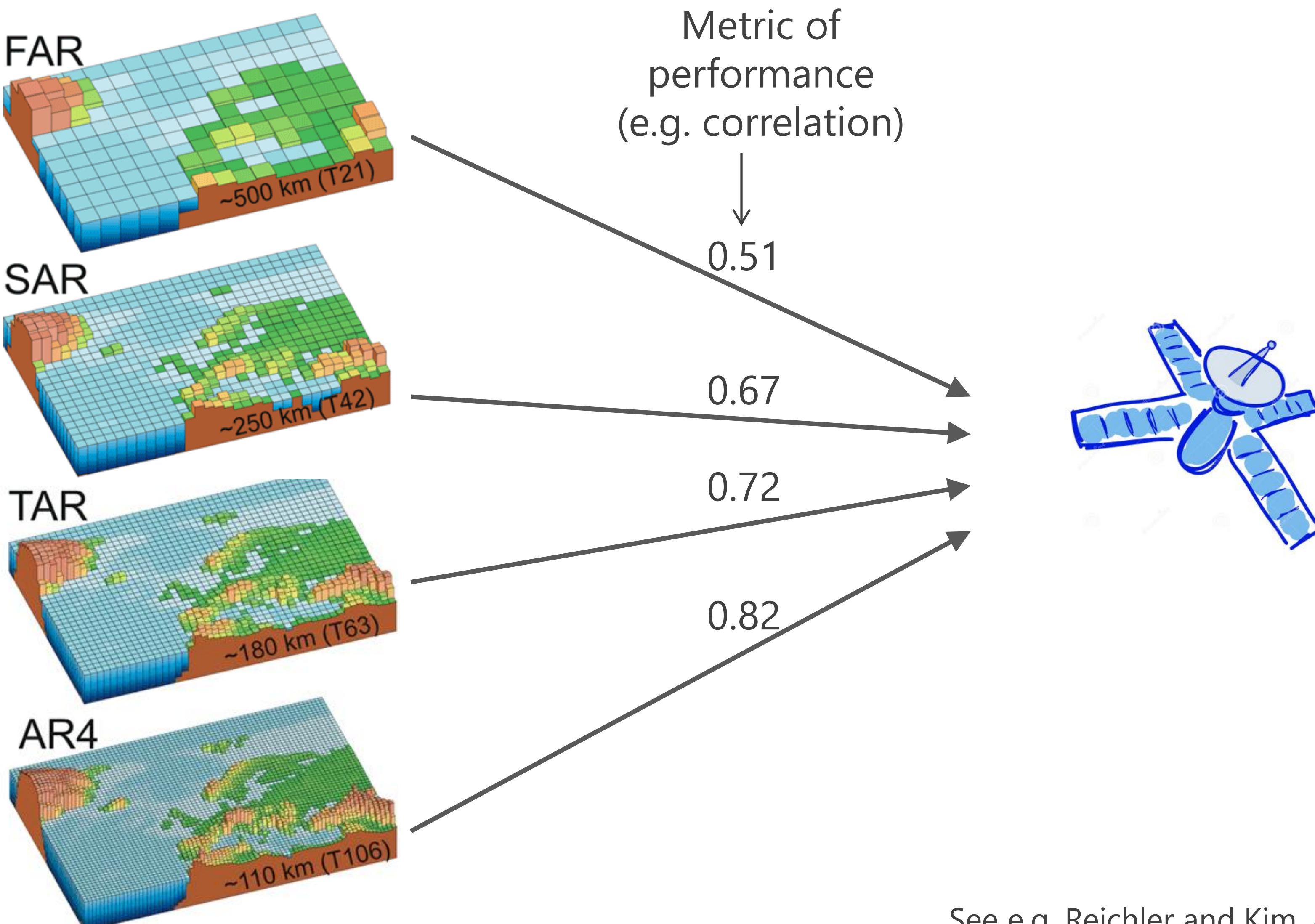
Sea ice observations are important for forecast verification
e.g., Sea Ice Prediction Network (SIPN) Sea Ice Outlook







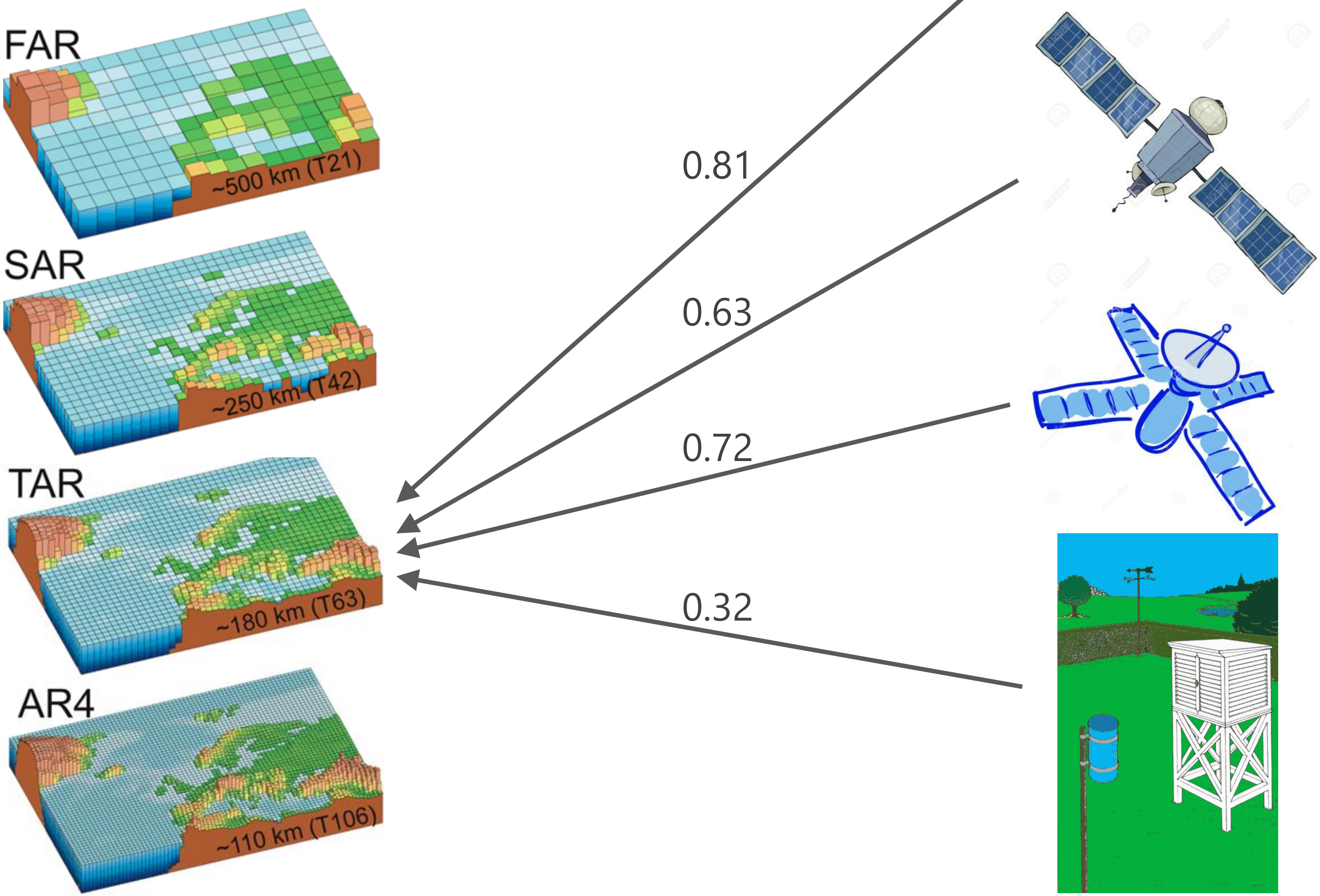
The classical approach of evaluation: several models for one observation



http://eo.ucar.edu/staff/rrussell/climate/modeling/climate_model_resolution.html

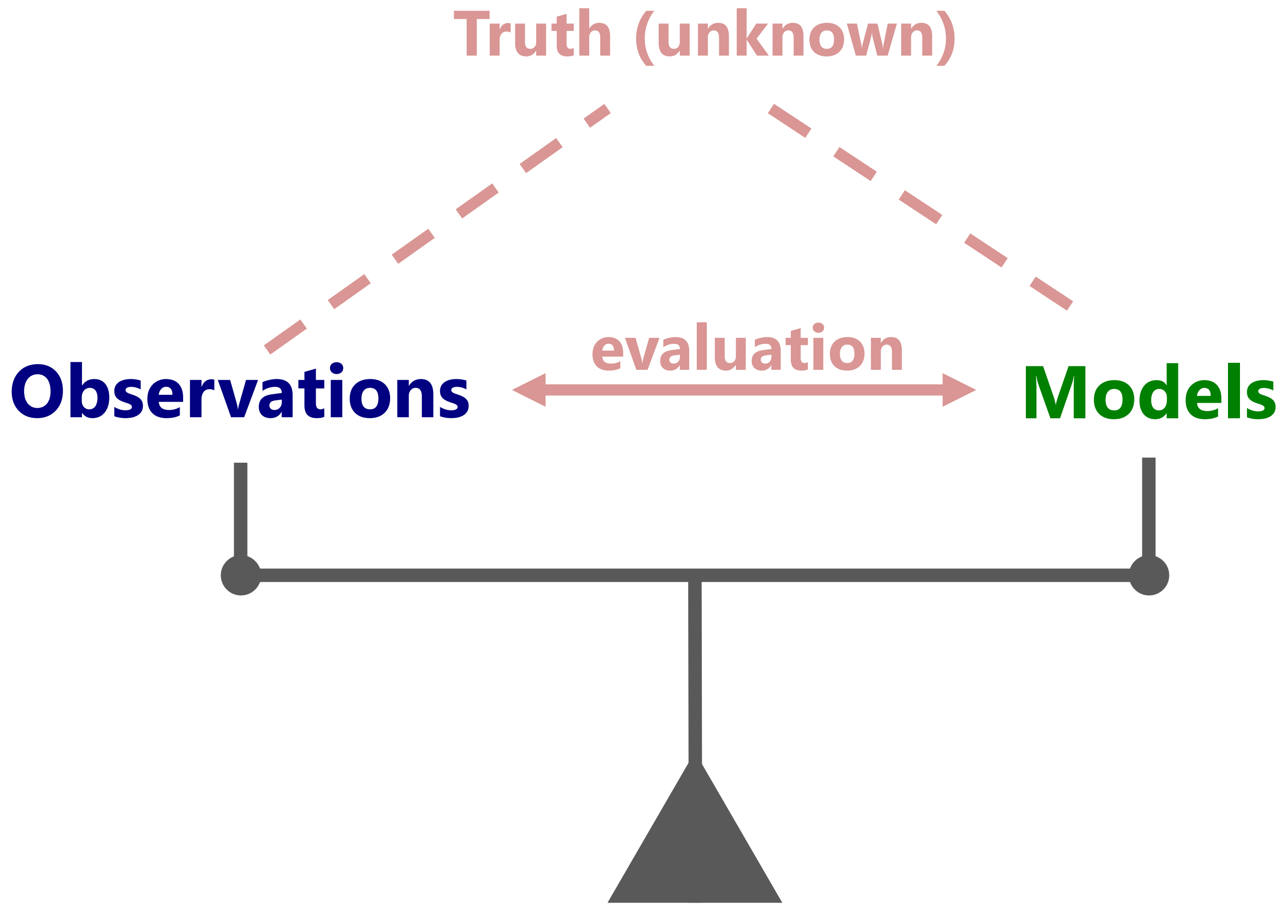
See e.g. Reichler and Kim, *BAMS*, 2008

Reversing the paradigm: several observations for one model



http://eo.ucar.edu/staff/rrussell/climate/modeling/climate_model_resolution.html

My understanding today



« Better observations yield better forecast verification scores »

1. Logic ▸ 2. Toy model ▸ 3. Climate models

« Better observations yield better forecast verification scores »

1. Logic ▸ 2. Toy model ▸ 3. Climate models

If metrics of performance (e.g., correlation, RMSE) are appropriate tools to reflect the quality of a modelling/forecast system,

Reichler and Kim, *BAMS*, 2008 (CMIP3→CMIP5)
Scaife et al., *GRL*, 2014 (NAO; sampling)
Massonnet et al., *The Cryosph*., 2012 (sea ice)
Msadek et al., *GRL*, 2014 (sea ice)

then, they can also reveal the underlying quality of an observational dataset.

This is because metrics of performance are symmetric from a mathematical point of view.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

« Better observations yield better forecast verification scores »

1. Logic ▸ 2. Toy model ▸ 3. Climate models

$\boxed{COR(A, B) = COR(B, A)}$

« Better observations yield better forecast verification scores »

1. Logic ▸ **2. Toy model** ▸ 3. Climate models

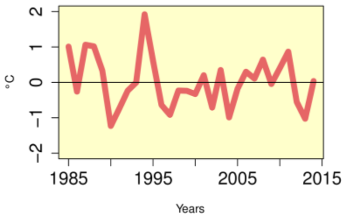
$\boxed{COR(A, B) = COR(B, A)}$

A signal-plus-noise toy model

Interannual variability

TRUTH $X_t = \boxed{\epsilon}$

$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$$

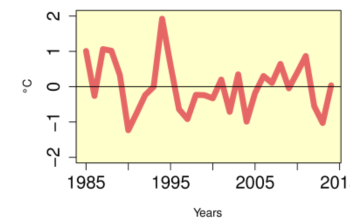


[Model from Weigel et al., *QJRMS*, 2008. See also Siegert et al., 2015]

A signal-plus-noise toy model

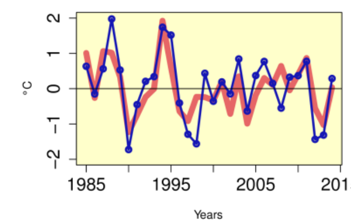
Interannual variability

TRUTH $X_t = \boxed{\epsilon}$ $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$



Measurement and representativity error

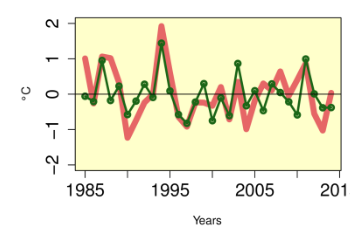
OBS $X_o = \epsilon + \boxed{\eta_o}$ $\eta_o \sim \mathcal{N}(0, \sigma_o)$



Model forecast error (physics, initial conditions, resolution) + irreducible error (atmosphere)

MODEL $X_f = \alpha\epsilon + \boxed{\eta_f + \eta_m}$ $\eta_f \sim \mathcal{N}(0, \sigma_f)$, $\eta_m \sim \mathcal{N}(0, \sigma_m)$

All error terms are assumed to be uncorrelated



[Model from Weigel et al., *QJRMS*, 2008. See also Siegert et al., 2015]

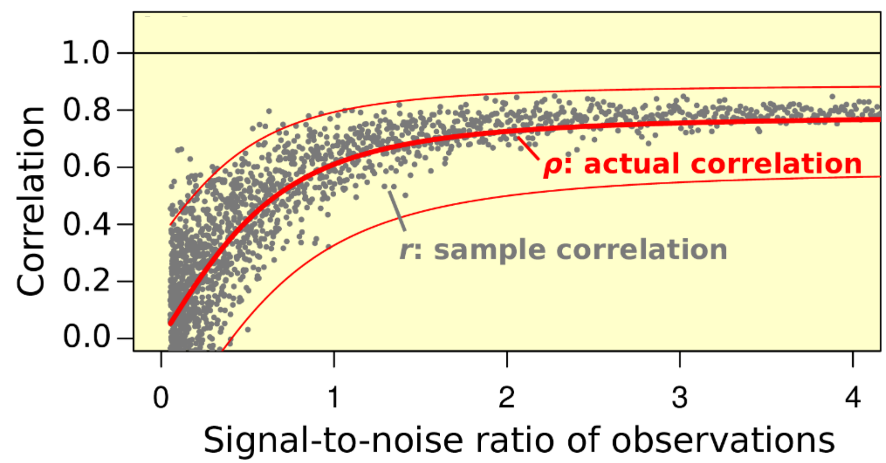
In this very simple paradigm, model and observational errors play interchangeable roles

$$\rho(X_o, X_f) = \frac{1}{\sqrt{\left(1 + \frac{\sigma_o^2}{\sigma_\epsilon^2}\right) \cdot \left(1 + \frac{(\sigma_f^2 + \sigma_m^2)/\alpha^2}{\sigma_\epsilon^2}\right)}}$$

Correlation **increases** when

- Model explains more variability,
- Model error decreases,
- Climate signal is stronger,
- Observational error decreases.

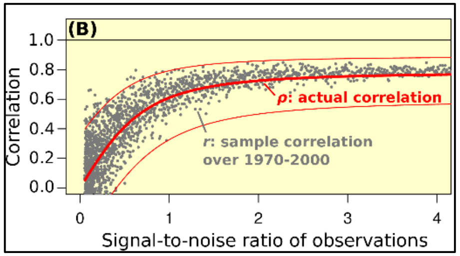
If error statistics are known, the dependence can be predicted



« Better observations yield better forecast verification scores »

1. Logic ▸ **2. Toy model** ▸ 3. Climate models

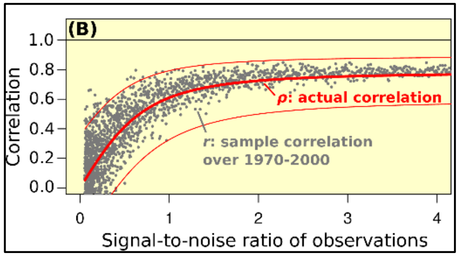
COR(A, B) = COR(B, A)



« Better observations yield better forecast verification scores »

1. Logic ▸ 2. Toy model ▸ **3. Climate models**

COR(A, B) = COR(B, A)



Seasonal forecasts of summer Arctic sea ice extent

CanCM3/4, MPI-ESM-L/M/HR, CNRM, EC-EARTH (3 versions)

9 models

10 members

4-month forecasts initialized in May

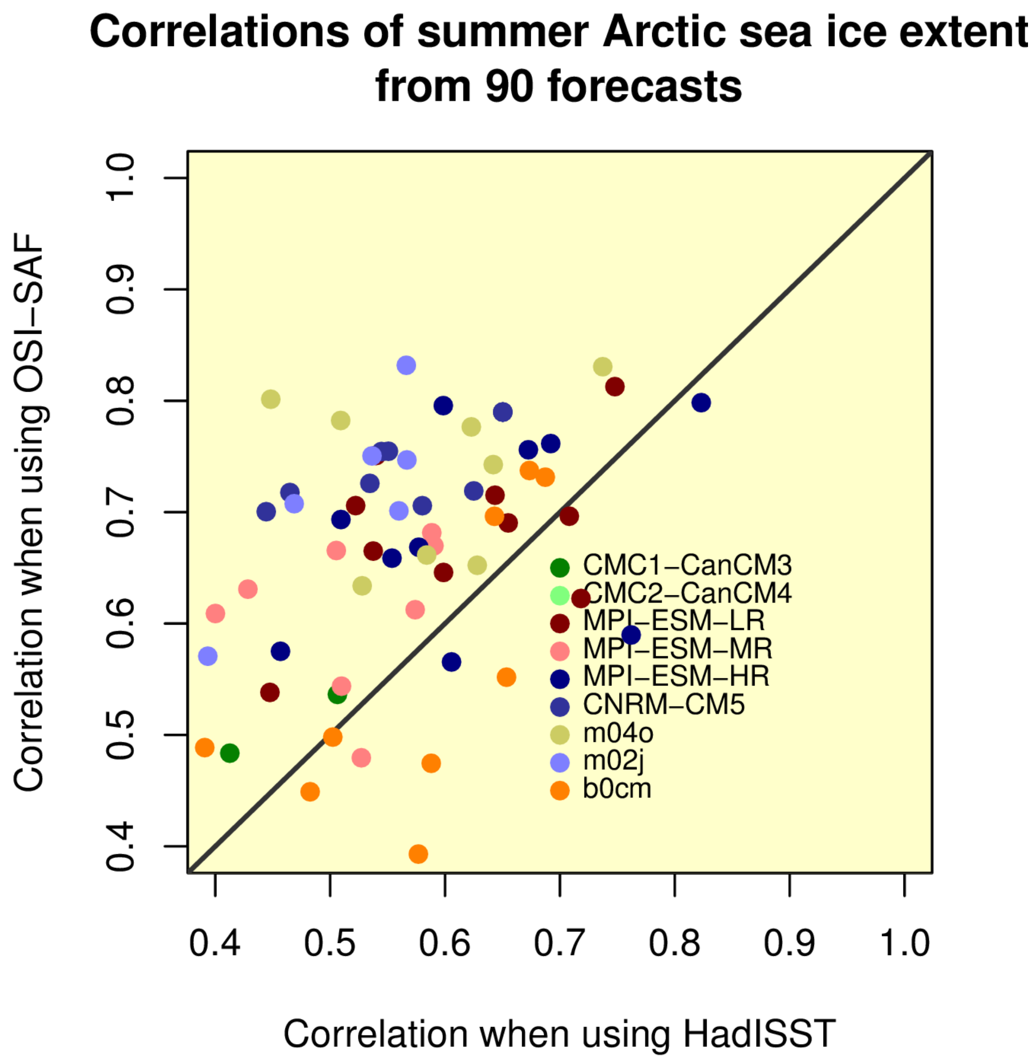
360 forecasts (not independent from each other)

Period for evaluation: 1993-2008

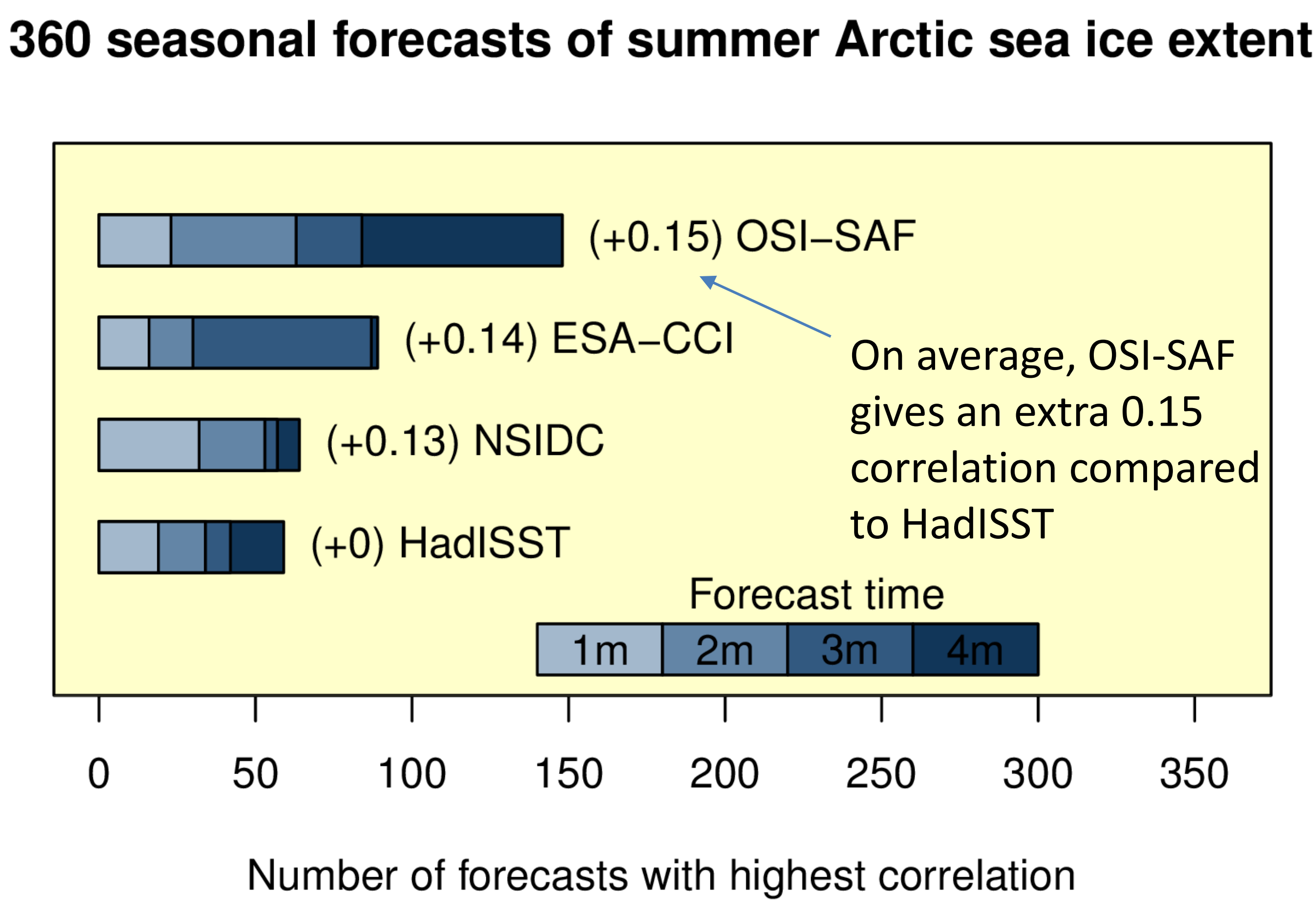
ESA-CCI OSI-SAF HadISST NSIDC

4 observational datasets for verification

(not independent from each other either)


Correlations of summer Arctic sea ice extent
from 90 forecasts
Correlation when using OSI−SAF
Correlation when using HadISST
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CMC1−CanCM3
CMC2−CanCM4
MPI−ESM−LR
MPI−ESM−MR
MPI−ESM−HR
CNRM−CM5
m04o
m02j
b0cm

Systematic dependence of skill score on the choice of verification product



Such an extreme result is unlikely to have occurred by chance

1. Bootstrapping

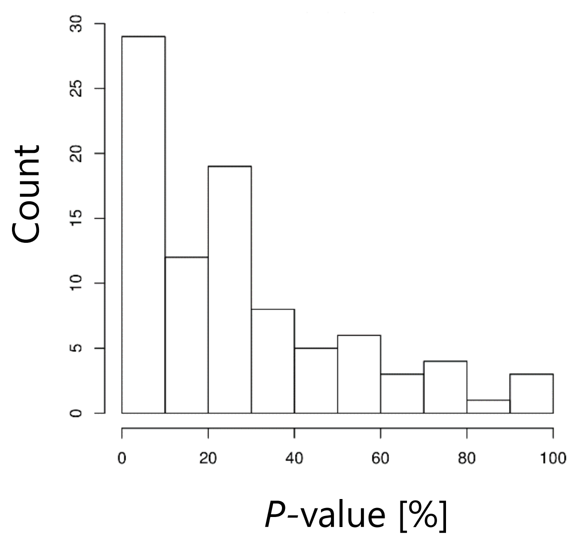
- Synthetic data is generated from the known sample covariance matrix of the data that we modify so that, for each forecast, correlations are set the same for all observations (our null hypothesis)

- With 10,000 trials, a result as extreme as the one we have happens ~0.2% of the time

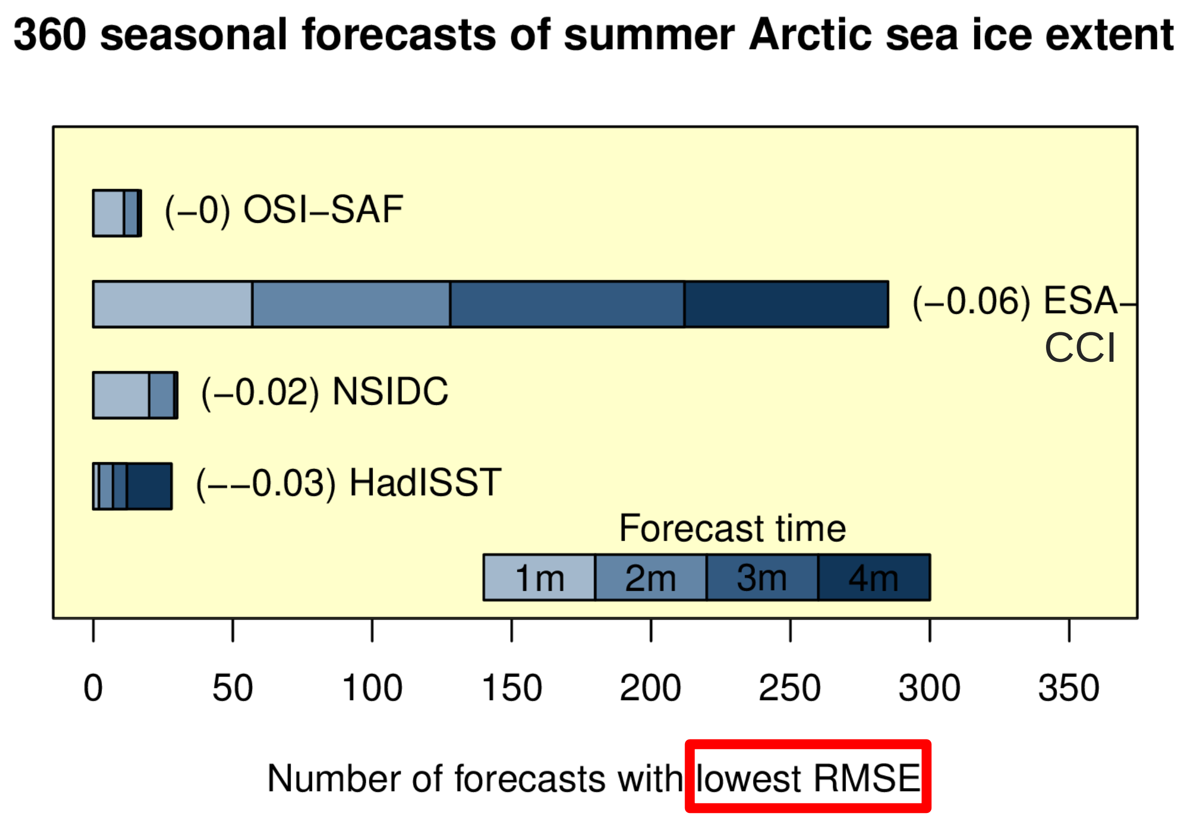
2. Parametric test

[Steiger et al., 1981]
(the test detects changes in correlation in presence of non-independent samples)

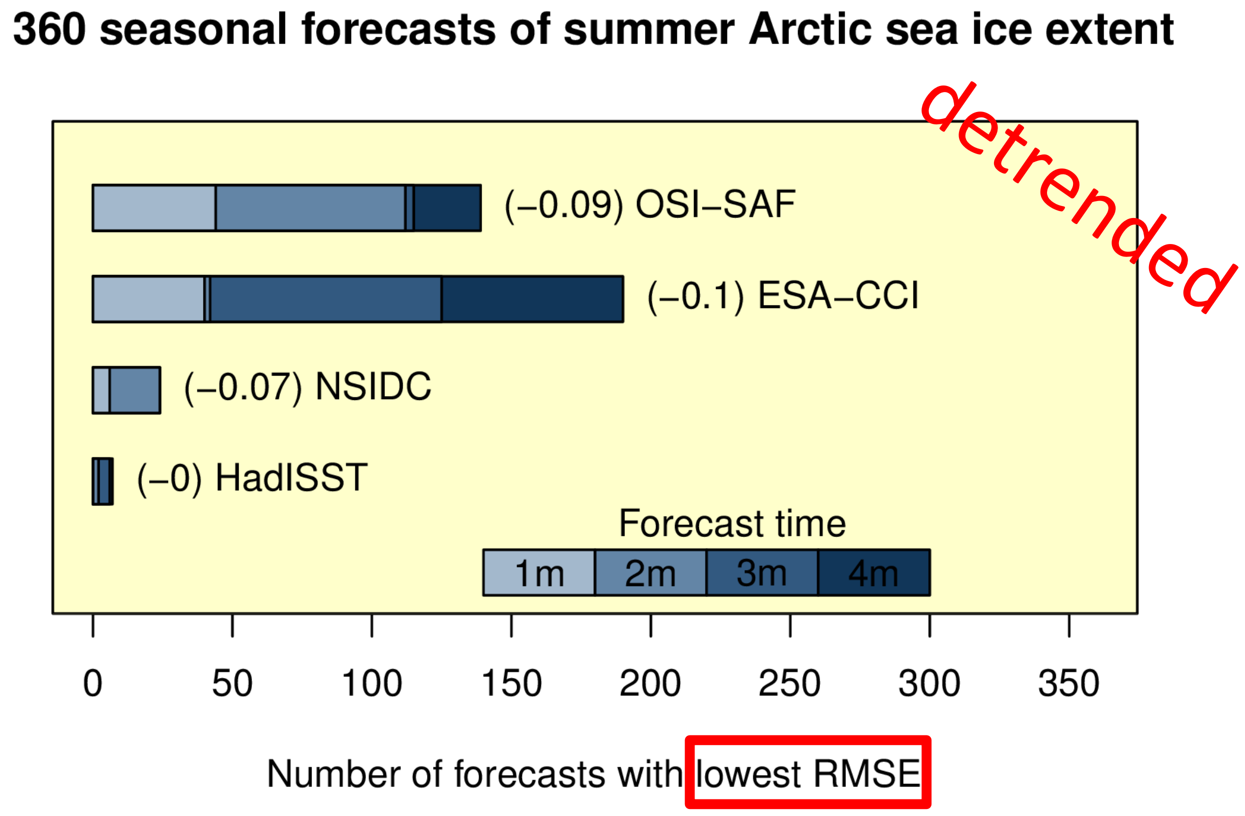
p-value of 110 Steiger tests to detect increase of correlation from HadISST (lowest) to OSI-SAF (highest)

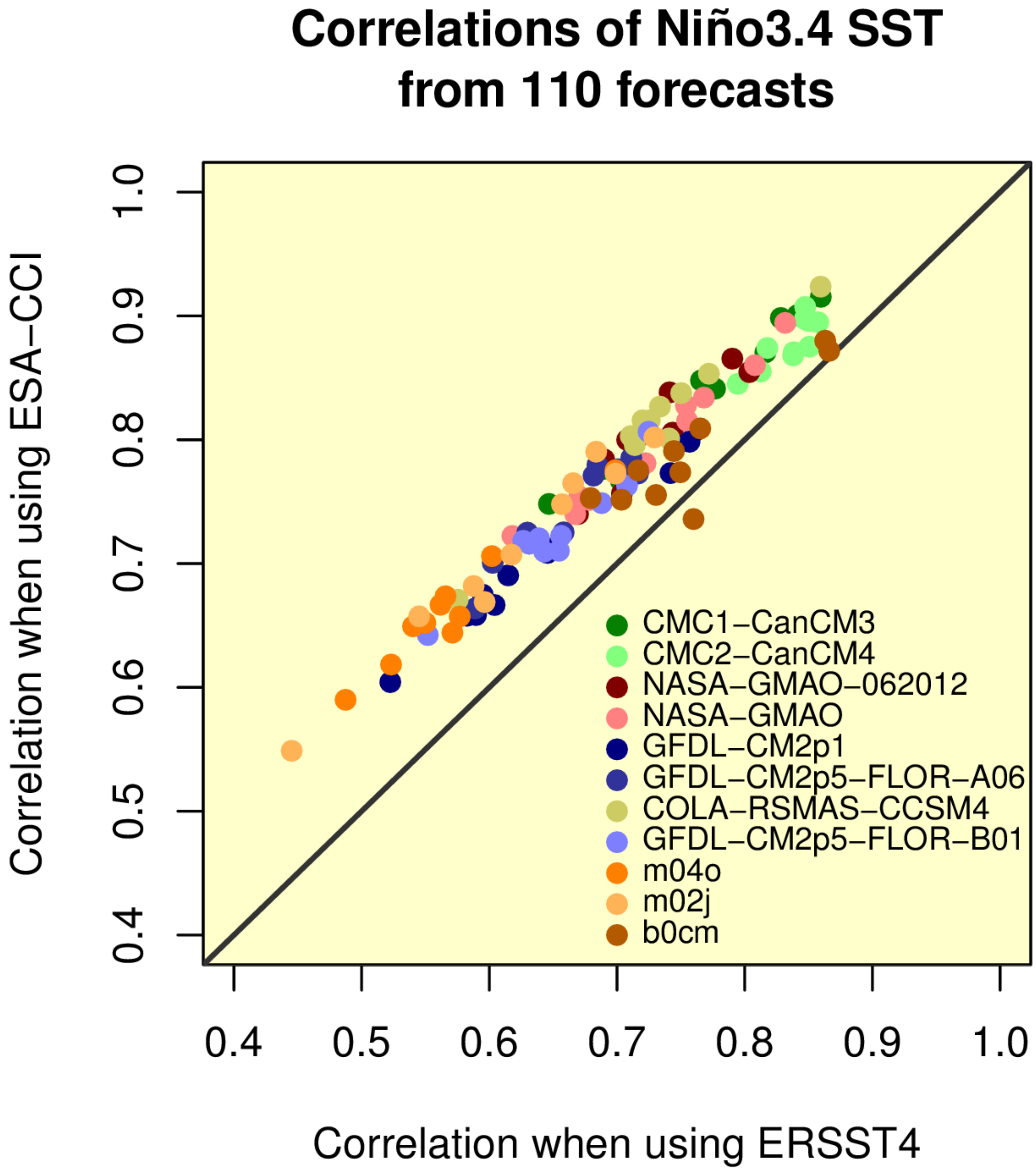


Systematic dependence of skill scores on verification product

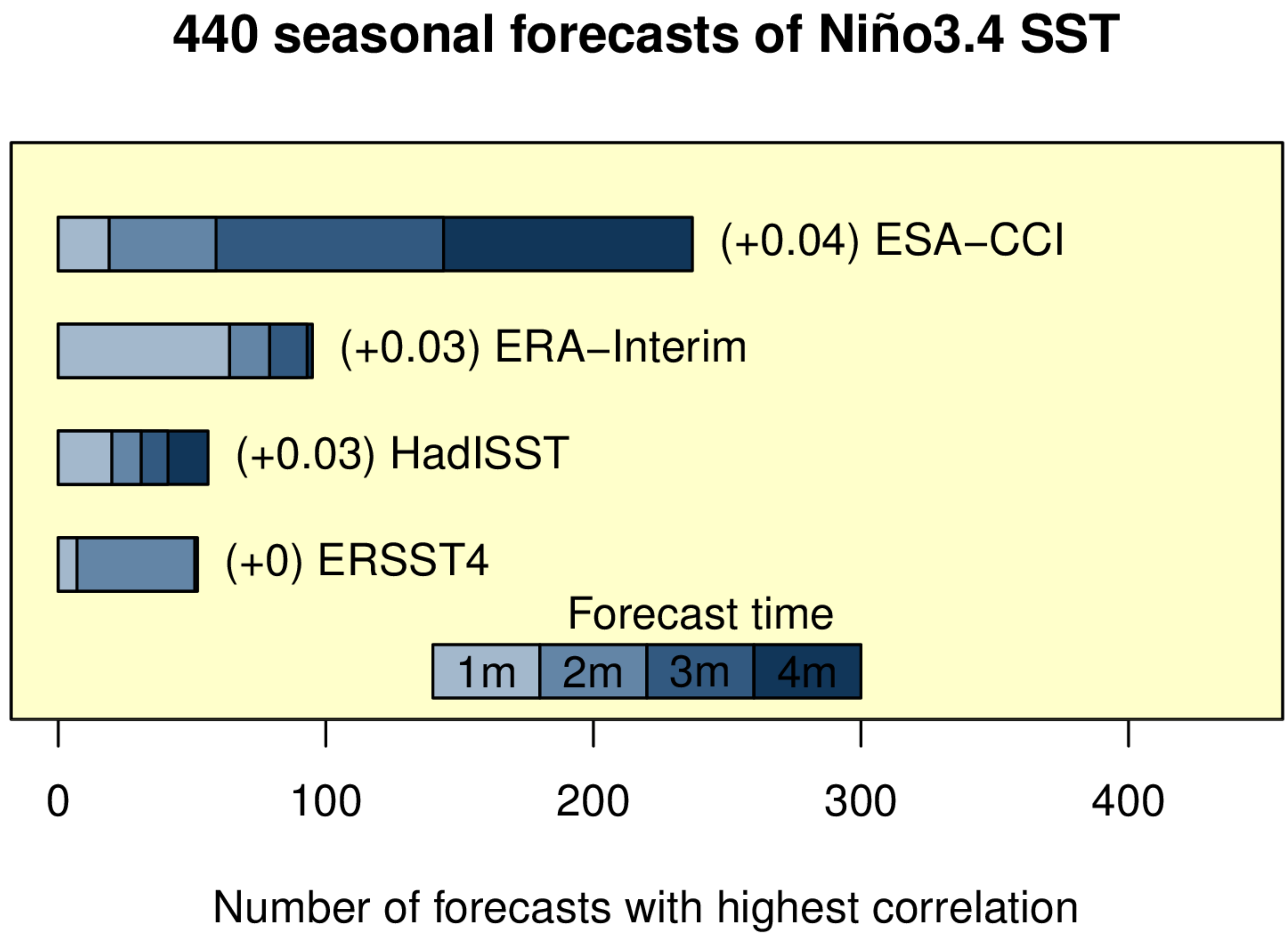


Systematic dependence of skill scores on verification product





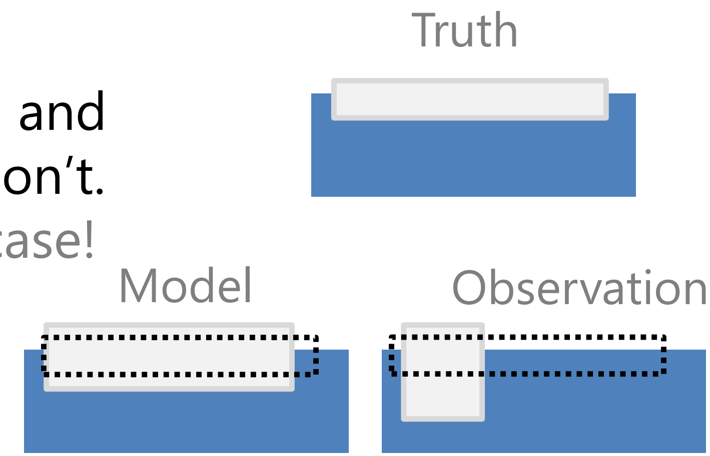
Correlations of Niño3.4 SST
from 110 forecasts
Correlation when using ESA−CCI
Correlation when using ERSST4
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CMC1−CanCM3
CMC2−CanCM4
NASA−GMAO−062012
NASA−GMAO
GFDL−CM2p1
GFDL−CM2p5−FLOR−A06
COLA−RSMAS−CCSM4
GFDL−CM2p5−FLOR−B01
m04o
m02j
b0cm



440 seasonal forecasts of Niño3.4 SST
(+0.04) ESA−CCI
(+0.03) ERA−Interim
(+0.03) HadISST
(+0) ERSST4
Forecast time
1m
2m
3m
4m
0
100
200
300
400
Number of forecasts with highest correlation

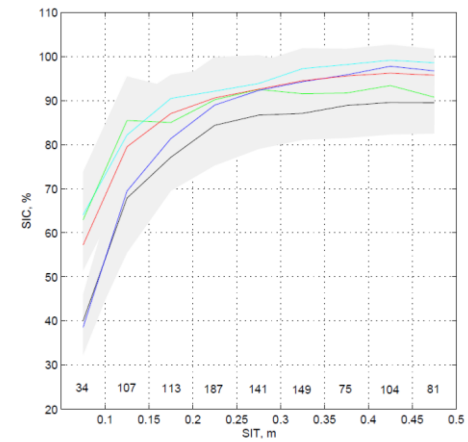
Why do models score better for the two most advanced and recent products?

Models simulate directly sea ice concentration and output it as a physical variable; observations don’t. Models can be really good references in that case!



Observations have deficiencies that models don’t have e.g. concentration of thin ice

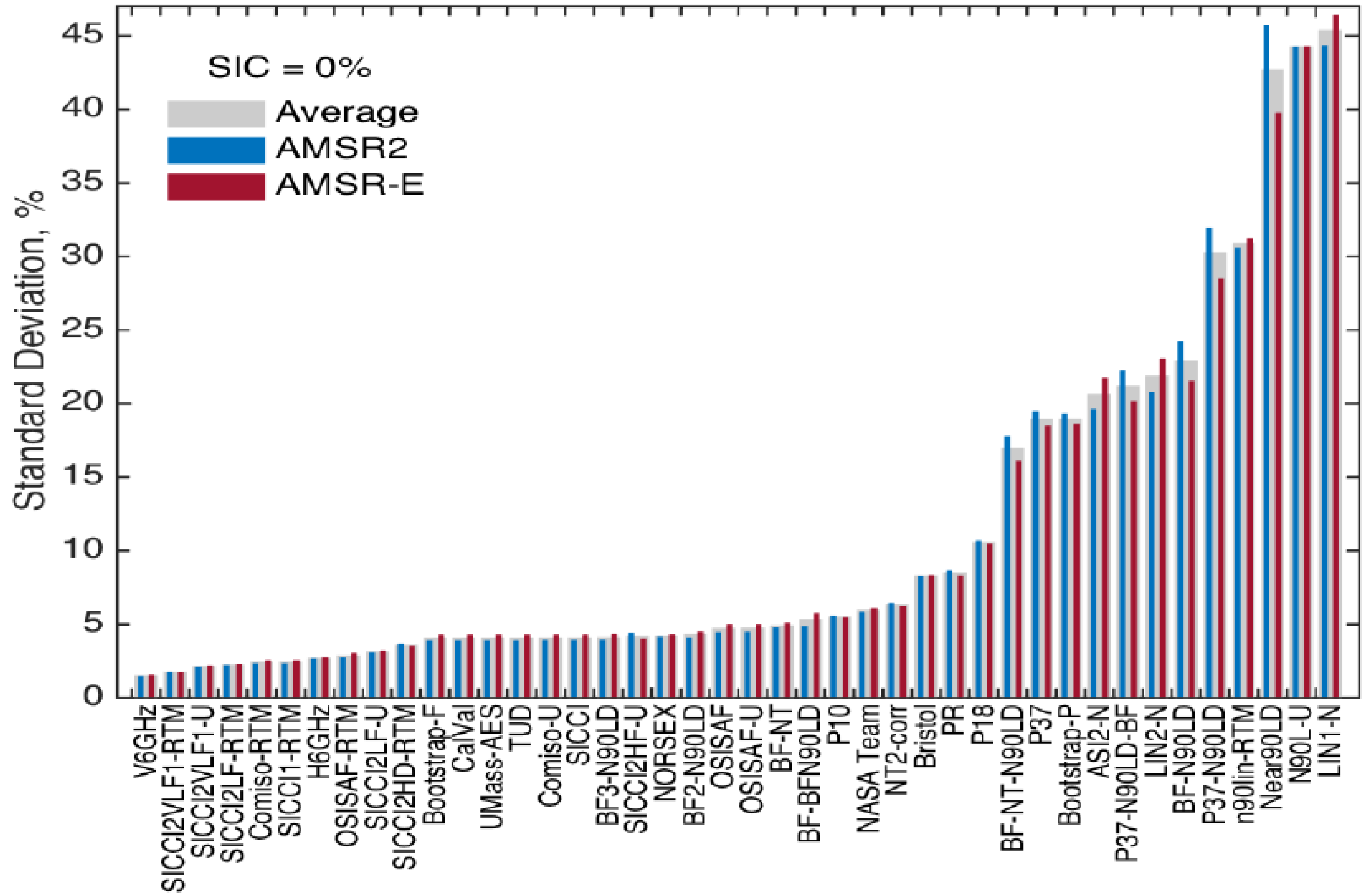
According to the toy model results, ESA-CCI and OSI-SAF should have lower errors (but only these two provide errors)



Note: remarkably, the models are also the most independent w.r.t. OSI-SAF and ESA-CCI

Sea ice concentration observational references	ESA-CCI	OSI-SAF	HadISST	NSIDC
Reference	Ivanova et al., 2015 (35)	Eastwood et al., 2015 (36)	Rayner et al., 2003 (33)	Fetterer and Knowles, 2004 (37)
Institution	European Space Agency	EUMETSAT	MetOffice (UK)	National Snow and Ice Data Center
Webpage	http://esa-cci.nersc.no/	http://osisaf.met.no/p/ice/ice_conc_reprocessed.html	http://www.metoffice.gov.uk/hadobs/hadisst/	https://nsidc.org/data/docs/noaa/g02135_seaice_index/
Period of product availability	1993-2008	1978-2015	1871-present	1978-present
Grid resolution	25 km x 25 km	10 km x 10 km	1.0°x1.0° (~110 km x 110 km near equator)	25 km x 25 km
Grid type	Equal-Area Scalable Earth	Polar Stereographic	Regular	Polar Stereographic
Primary product and technology on which the analysis is based	Passive microwave satellite data: SSM/I (1992-2008)	Passive microwave satellite data: SMMR and SSM/I-SSMIS.	Passive microwave satellite data: SMMR and SSM/I-SSMIS (1978-1996) + NCEP operational dataset (1997-present)	Passive microwave satellite data: SMMR and SSM/I-SSMIS.
Algorithm of processing	Same as OSI-SAF algorithm, but improved so that better performance is achieved over thin ice. Atmospheric filter was applied to brightness temperature directly, and not on the spectral gradient ratio as this latter approach was found to eliminate low ice concentrations.	Hybrid: Bootstrap and Bristol. **Dynamical tie-points** are used (calibration parameters are time-dependent). Weather filter was improved from NASA Team algorithm, as it was found that this correction tended to eliminate low ice concentrations.	NASA Team (see NSIDC column)	NASA Team: **Static tie-points** are used, but different datasets for Northren and Southern Hemispheres. Weather filter was used: SIC set to zero when the spectral gradient ratio (GR) is > 0.07
Other comments	Only ESA-CCI phase 1 products are considered. Phase 2 products will be developed joinly with EUMETSAT OSI-SAF.	Version 1.2 of the product was used, i.e. without input data from ESA. The next release of OSI-SAF reprocessed sea ice concentration will include results from the ESA-CCI research project	The HadISST product is updated with NCEP operational analyses from 1997 onwards.	We take as monthly sea ice extent the value already processed by NSIDC.

ICE CONCENTRATION ALGORITHM PERFORMANCE (SIC=0)

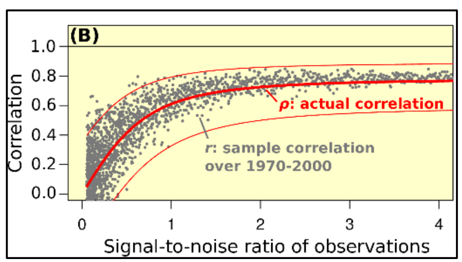


Courtesy L. T. Pedersen,
N. Ivanova and co-authors

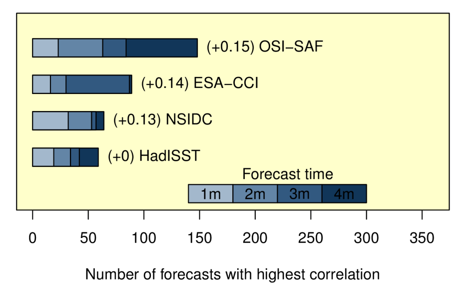
« Better observations yield better forecast verification scores »

1. Logic ▸ 2. Toy model ▸ **3. Climate models**

$$COR(A, B) = COR(B, A)$$



360 seasonal forecasts of summer Arctic sea ice extent



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Conclusions & Outlooks

Summary | Using three types of arguments (intuitive reasoning, idealized model, real forecasts) we find evidence that better observations increase forecast skill as we estimate it routinely in the community of seasonal forecasting.

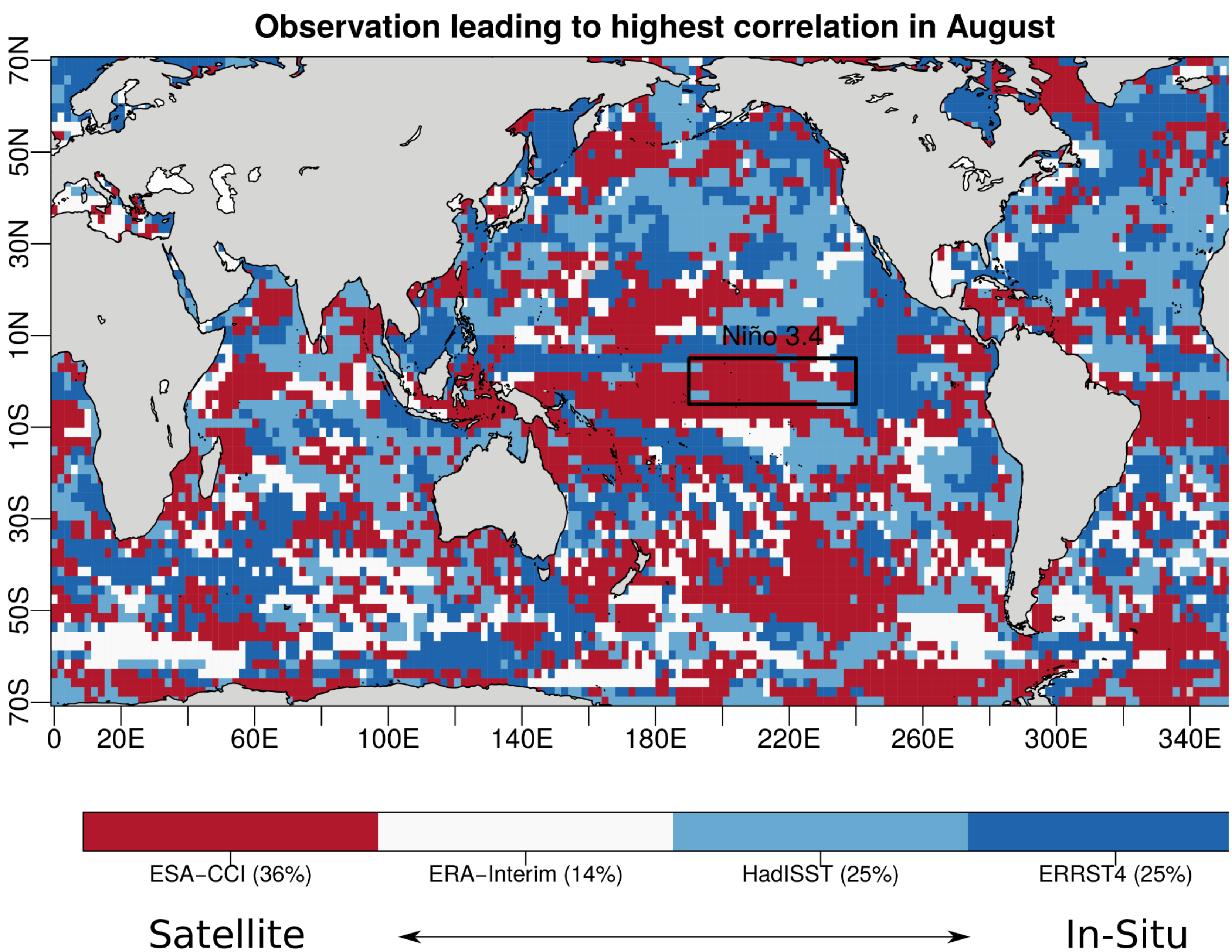
Interpretation | These results are best understood if observations and models are considered at the same level (i.e., observations are not superior to models). Observational errors will systematically lower actual forecast skill, in the same way that model errors systematically lower forecast skill.

Recommendations | Models should not always be blamed for low performance. Observations can do this job easily! This overlooked source of error can give differences as large as differences from one model version to another. Modellers should be careful in picking their observational product, or at least use several of them.

Outlooks | Quantifying observational error propagation over time-averaged periods, space-averaged domains, is key to introduce observational uncertainty in current metrics of performance. Yet these error statistics depend on many unknowns, such as decorrelation time- and space-scales between grid-point, daily error statistics.

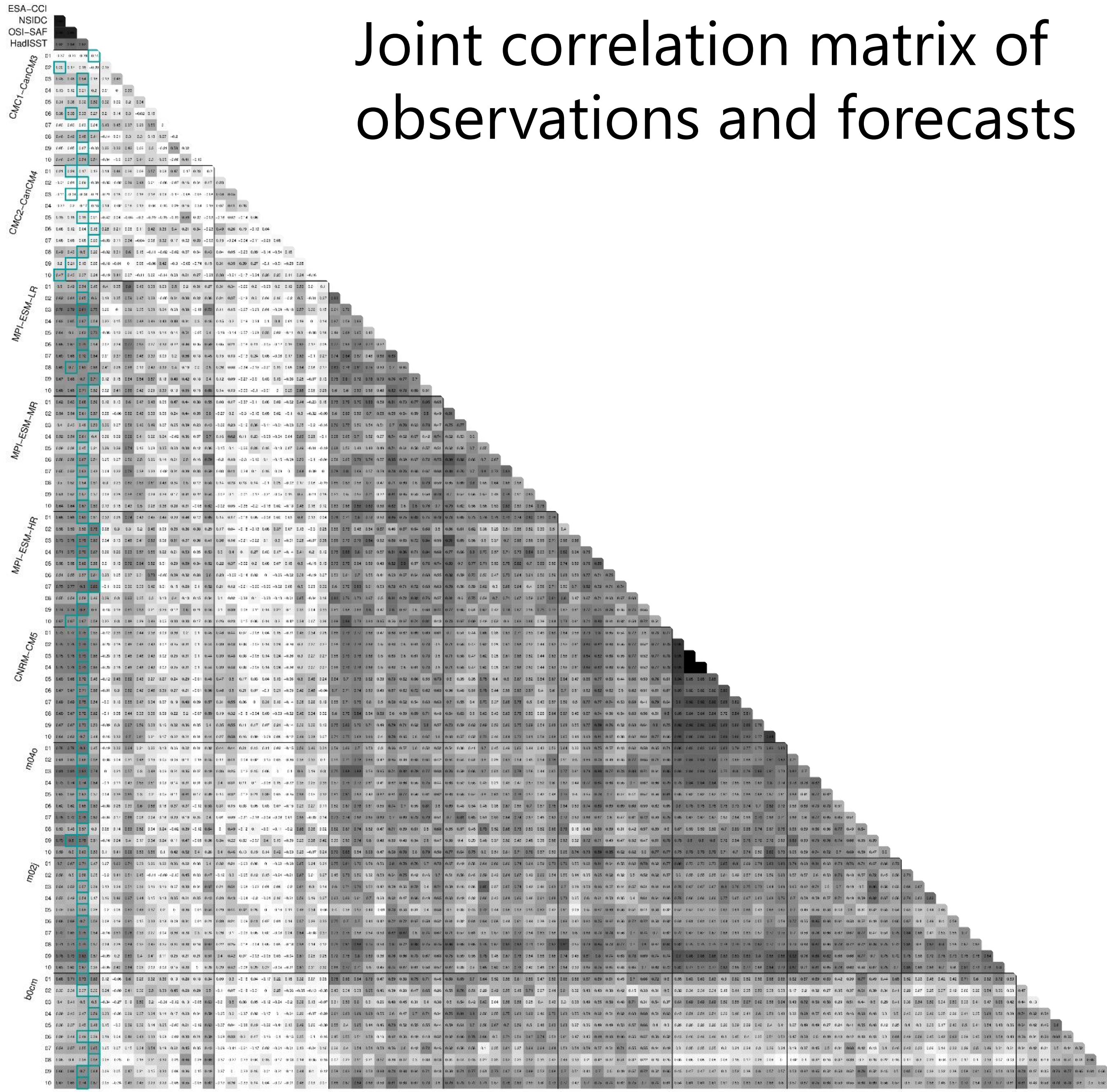
Thank you!

francois.massonnet@bsc.es


Observation leading to highest correlation in August
70N
50N
30N
10N
10S
30S
50S
70S
0
20E
60E
100E
140E
180E
220E
260E
300E
340E
Niño 3.4
ESA−CCI (36%)
ERA−Interim (14%)
HadISST (25%)
ERRST4 (25%)
Satellite
In-Situ

Joint correlation matrix of observations and forecasts

Correlation of August sea ice extent (1993-2008)



Original covariance matrix

	OBS 1	OBS 2	FORECAST 1	FORECAST 2	FORECAST 3
OBS 1	σ_{o1}^2				
OBS 2	$r_{o1,o2} \,.\, s_{o1} .\, s_{o2}$	σ_{o2}^2			
FORECAST 1	$r_{o1,f1} \,.\, \sigma_{o1} .\, \sigma_{f1}$	$r_{o2,f1} \,.\, \sigma_{o2} .\, \sigma_{f1}$	σ_{f1}^2		
FORECAST 2	$r_{o1,f2} \,.\, \sigma_{o1} .\, \sigma_{f2}$	$r_{o2,f2} \,.\, \sigma_{o2} .\, \sigma_{f2}$	$r_{f1,f2} \,.\, \sigma_{f1} .\, \sigma_{f2}$	σ_{f2}^2	
FORECAST 3	$r_{o1,f3} \,.\, \sigma_{o1} .\, \sigma_{f3}$	$r_{o2,f3} \,.\, \sigma_{o2} .\, \sigma_{f3}$	$r_{f1,f3} \,.\, \sigma_{f1} .\, \sigma_{f3}$	$r_{f2,f3} \,.\, \sigma_{f2} .\, \sigma_{f3}$	σ_{f3}^2

Modified covariance matrix

	OBS 1	OBS 2	FORECAST 1	FORECAST 2	FORECAST 3
OBS 1	σ_{o1}^2				
OBS 2	$r_{o1,o2} \,.\, \sigma_{o1} .\, \sigma_{o2}$	σ_{o2}^2			
FORECAST 1	$r_{1} \,.\, \sigma_{o1} .\, \sigma_{f1}$	$r_{1} \,.\, \sigma_{o2} .\, \sigma_{f1}$	σ_{f1}^2		
FORECAST 2	$r_{2} \,.\, \sigma_{o1} .\, \sigma_{f2}$	$r_{2} \,.\, \sigma_{o2} .\, \sigma_{f2}$	$r_{f1,f2} \,.\, \sigma_{f1} .\, \sigma_{f2}$	σ_{f2}^2	
FORECAST 3	$r_{3} \,.\, \sigma_{o1} .\, \sigma_{f3}$	$r_{3} \,.\, \sigma_{o2} .\, \sigma_{f3}$	$r_{f1,f3} \,.\, \sigma_{f1} .\, \sigma_{f3}$	$r_{f2,f3} \,.\, \sigma_{f2} .\, \sigma_{f3}$	σ_{f3}^2